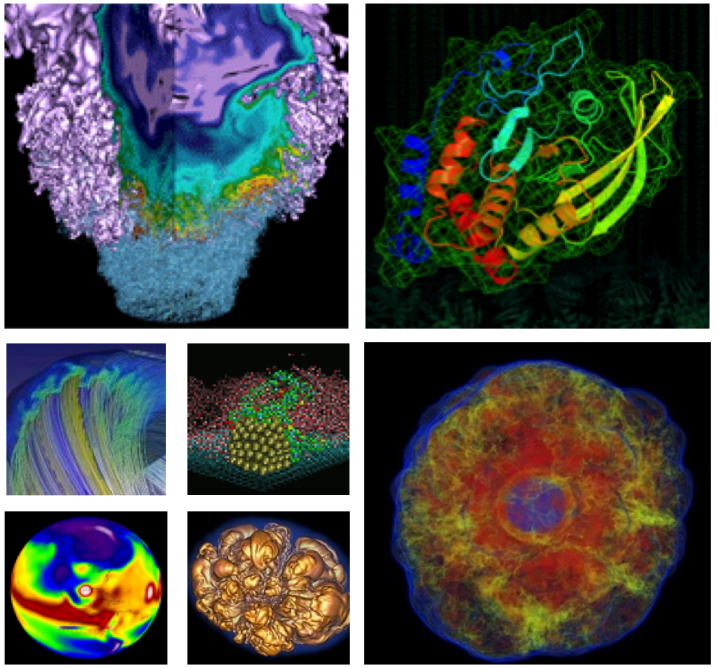
With Big Data Comes Big Responsibility

Operational Data Analytics
at the National Energy Research Scientific Computing Center



NERSC

Melissa Romanus

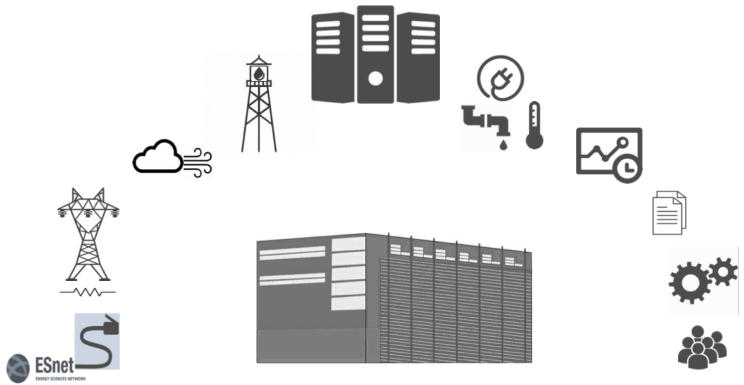
SC'19 – Operational Data Analytics BOF
November 21, 2019

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Key Characteristics of HPC DC OD

- Heterogeneous sources
- Heterogeneous data fields & formats
- Distributed
- Continuous 24x7
  - Nonlinear time-series
  - Realtime streaming
- Combination of pure metric (e.g., network counters) and more complex text (e.g., syslogs)



NERSC HPC & Operational Data

Machine Metrics

Network Counters and Flow

Particle Sensors

Job Scheduling

Application Characteristics

Power Usage & Distribution

Node Status

Filesystem I/O

CPU Usage & Temperature

Air Circulation

Outdoor Weather Station

Syslogs

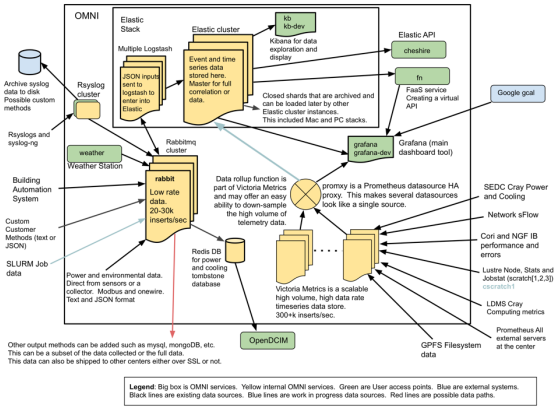
Water Tower Fans & Pump Status

Cooled Water Temperature & Pressure

Room, Rack, and Cabinet Temperature & Humidity

OMNI Design Principles & Use Cases

- Collect everything
- Keep it forever
- Accessible in near real-time
- High-availability
- Scalable to emerging systems
- Maintainable



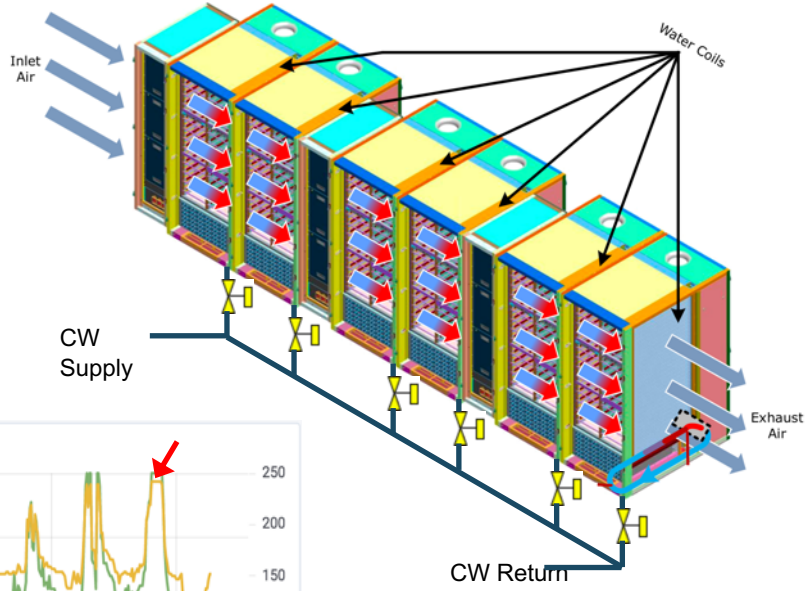
Source: Elizabeth Bautista, Melissa Romanus, et. al. 2019. *Collecting, Monitoring, and Analyzing Facility and Systems Data at the National Energy Research Scientific Computing Center.* In Proceedings of the 48th International Conference on Parallel Processing: Workshops (ICPP 2019). ACM, New York, NY, USA. https://doi.org/10.1145/3339186.3339213

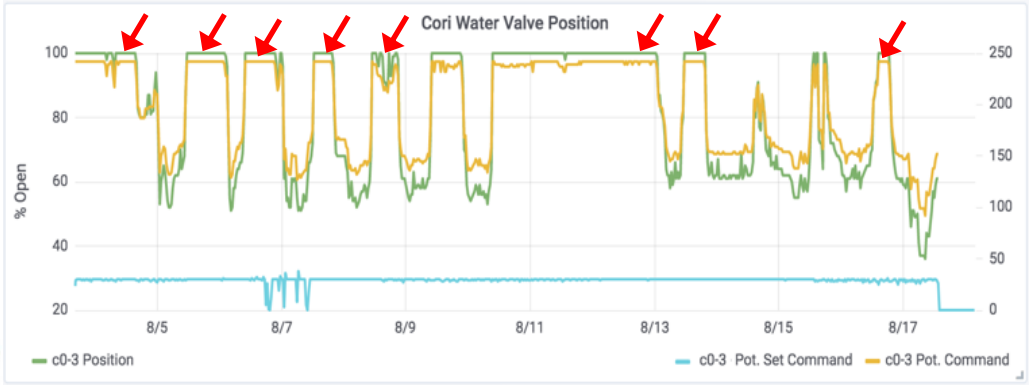
Now What?

Image: LaBan, Terry. *Comic Brake*, Field Service Digital, 3 Nov. 2017, https://www.servicemax.com/uk/fsd/comic-brake-thought-big-data-difficult-manage-meet-ginormous-data.

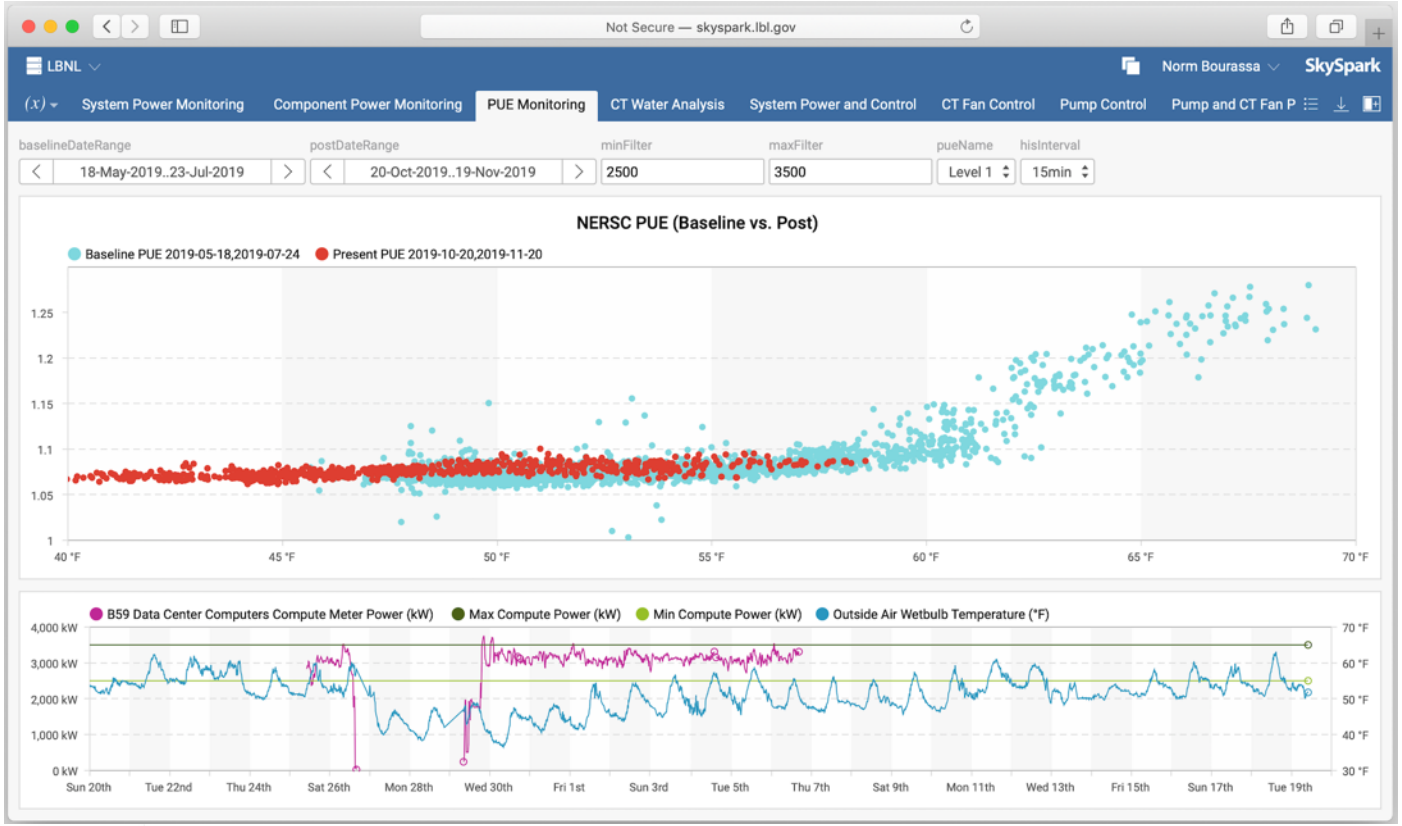
Merged System & Infrastructure Data

- **Out of the Box Operation:** *Static cabinet air temp setpoint*
  - Can't set for all weather, therefore CW pumping and fan energy waste
- **Cooling Plant Interactive Operation:** *Variable cabinet air temp setpoints*
  - **Cooler CW temps:** Cooler cabinet air therefore fan speed turndown
  - **Warmer CW Temps:** Fan speeds compensate. Dynamic setpoint reduces excessive CW demand valve positions





Quantifying Energy Efficiency Changes



What's Next?

- Vendor collaborations are essential for collecting necessary data and implementing persistent availability
  - How do we engage vendors more?
- Can we generalize solutions across HPC data centers? What data can be shared?
- How can we achieve automation or semi-automation given this data?
  - Capturing human expertise
  - "Experimenting" when the systems are worth millions of dollars
  - Changes need to take the entire data center into account

Thank You

Melissa Romanus
NERSC
Operations Technology Group
mromanus@lbl.gov